Geneva Workshop 2015

Critical appraisal of the medical literature

Jane Hirst MBBS, MPH, PhD FRANZCOG

University of Oxford

UNIVERSITY OF OXFORD

What makes a study reliable?

Nine out of ten women we asked believe that "Youth Code" makes their skin firmer and younger looking

- Which design would you choose to maximise the chance of getting the result you want?

a) Ask women buying "Youth Code" in the shops whether they agree their skin is firmer and younger looking?

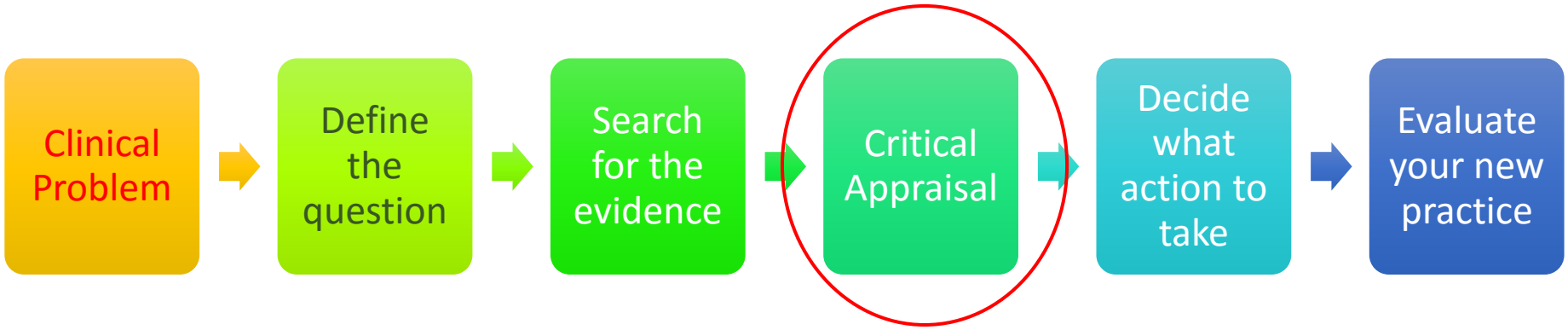
a) Ask a random sample of women to try "Youth Code" and then comment on whether they agree that their skin is firmer and younger looking?

What is critical appraisal?

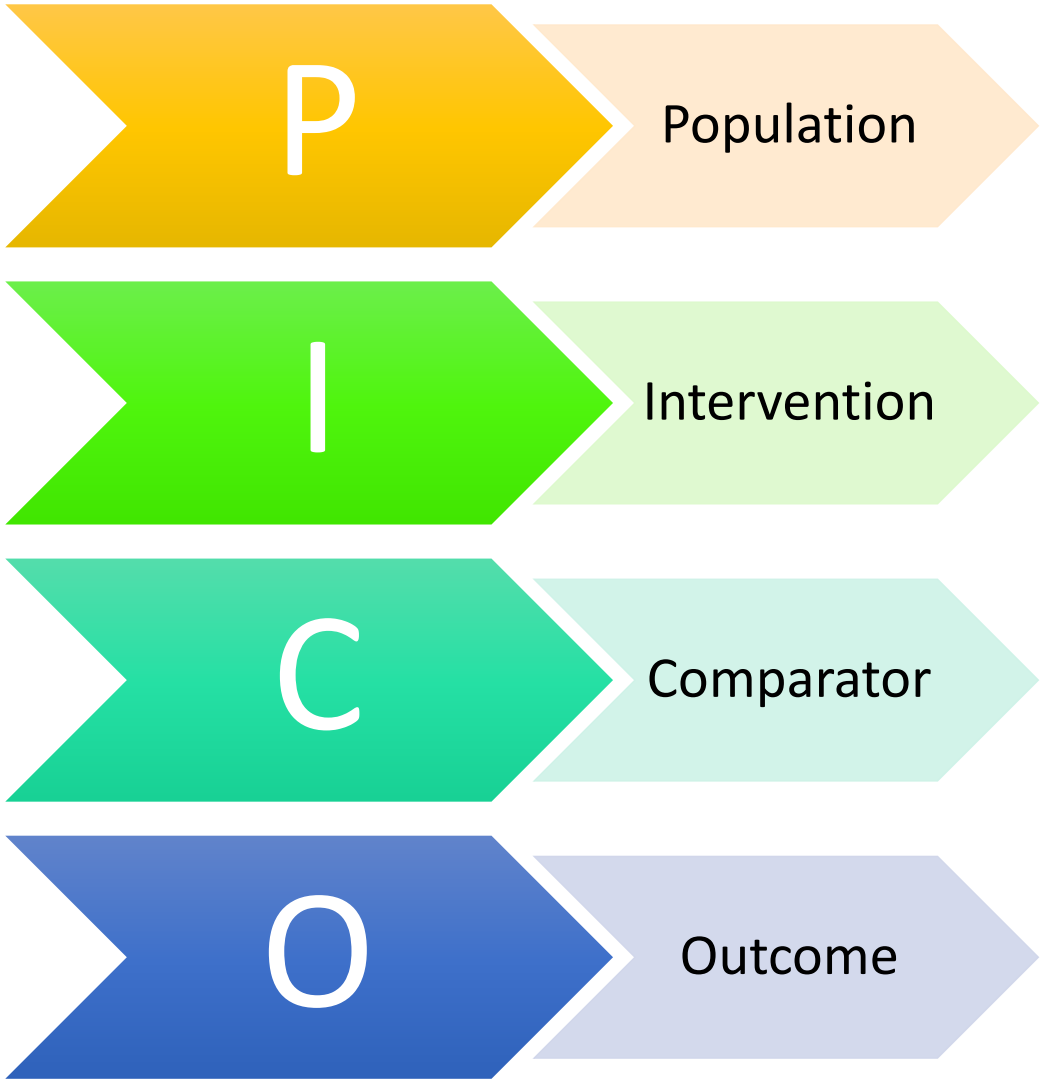
- Carefully and systematically evaluate research to assess:
  - Validity (is these findings trustworthy?)
  - Value (what do the results show?)
  - Relevance (How do these results relate to my clinical practice?)

Burls 2009

Critical appraisal: a key component of evidence based medicine



Asking the right question



	1	2	3	4
	Patient or Problem	Intervention (a cause, prognostic factor, treatment, etc.)	Comparison Intervention (if necessary)	Outcomes
Tips for Building	Starting with your patient, ask "How would I describe a group of patients similar to mine?" Balance precision with brevity.	Ask "Which main intervention am I considering?"Be specific.	Ask "What is the main alternative to compare with the intervention?"Again, be specific.	Ask "What can I hope to accomplish?" or "What could this exposure really affect?"Again, be specific.
Example	"In patients with heart failure from dilated cardiomyopathy who are in sinus rhythm …"	"… would adding anticoagulation with warfarin to standard heart failure therapy …"	"… when compared with standard therapy alone …"	"… lead to lower mortality or morbidity from thromboembolism. Is this enough to be worth the increased risk of bleeding?"

cebm.net

Choosing right study design

- Some study designs are not appropriate to answer certain questions

- All study designs are prone to different biases

Pyramid of evidence

So are RCTs the gold standard for evidence?

…..depends

Slides from: K Mahtani, CEBM Oxford

Limitations of RCTs

- Excellent vs Poor RCTs – quality varies
  - Impact on interpretation of result (external validity)?
- Expensive and time consuming
  - £250k - £millions over 2-5 years+
- May not always be the right study design to answer that question

A RCT to examine if smoking causes lung cancer

- 30 healthy Oxford Students
- Randomise to 2 groups
  - Gp1 smokes 20 cigarettes per day every day
  - Gp2 no smoking

wellcome trust

NHS National Institute for Health Research

MRC Medical Research Council

Types of research

- What is the best study design for answering this type of question?
  - **Aetiology**
  - **Diagnosis**
  - **Prognosis**
  - **Harm**
  - **Effectiveness**
  - **Qualitative**

How to critically appraise an article

- **Validity:** methods to check that the **biases** for which that particular study design is prone have been minimised
- **Results**
- **Clinical relevance**

Validity

Internal

External

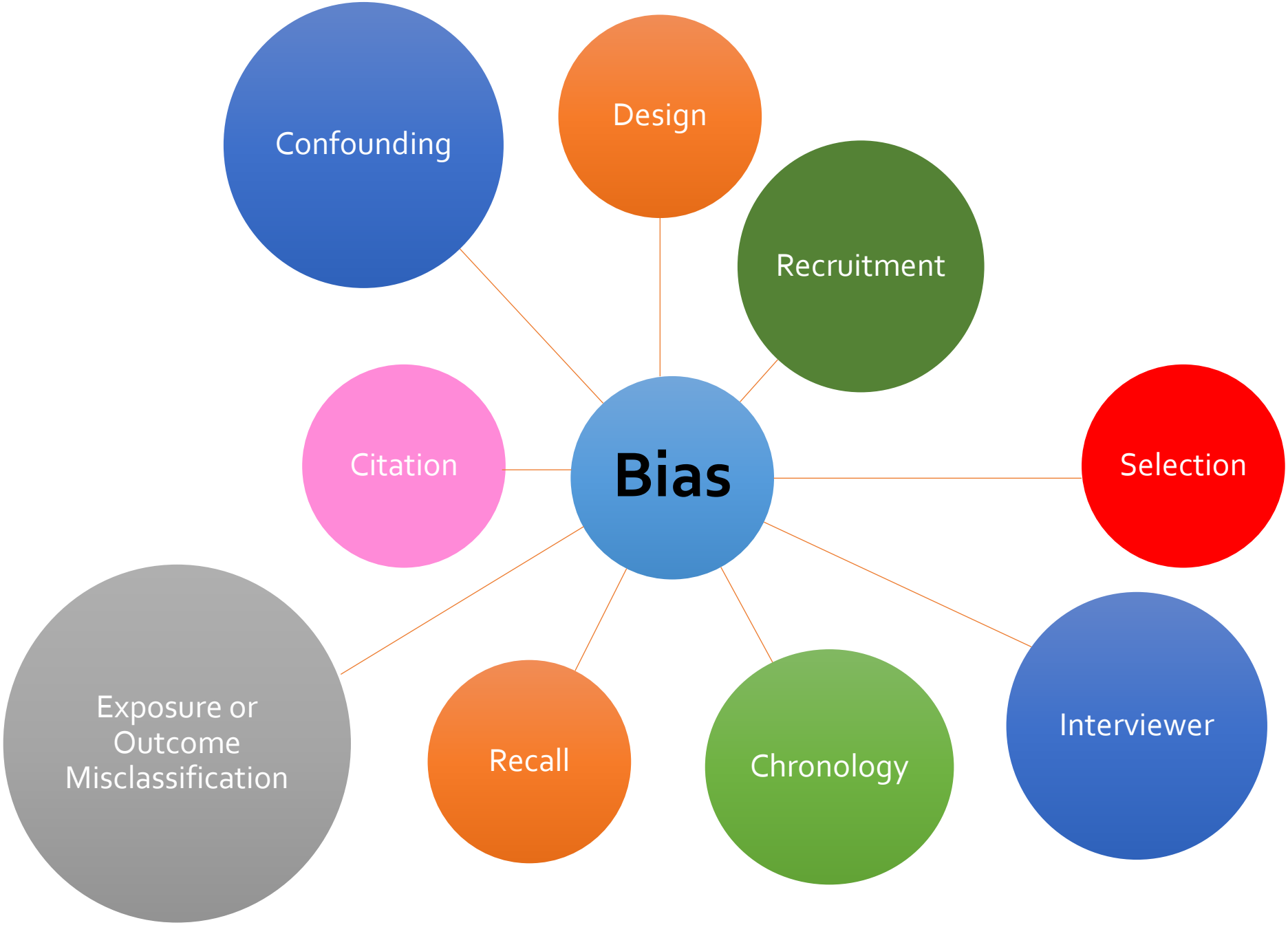
Bias

"the systematic deviation of the results of a study from the truth because of the way it has been conducted, analysed or reported"

Burls, "What is Critical Appraisal" 2009

Internal validity

- No study is perfect and completely free from bias
- Have the researchers done all they can to minimise bias?
- Are the biases that remain unlikely to have affected the final results?


Confounding
Design
Recruitment
Citation
Bias
Selection
Exposure or Outcome Misclassification
Recall
Chronology
Interviewer

Sources of bias in clinical trials

Table 1. Key sources of bias in clinical trials[2]

Selection bias	Biased allocation to comparison groups
Performance bias	Unequal provision of care apart from treatment under evaluation
Detection bias	Biased assessment of outcome
Attrition bias	Biased occurrence and handling of deviations from protocol and loss to follow up

Juni, BMJ 2001

Assessing Trials of effectiveness

Questions to ask:

1. Are the results of the trial valid?
2. What are the results?
3. Will the results help locally?

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net

Checklists for clinical trials

SIGN

CEBM

CENTRE FOR EVIDENCE-BASED MEDICINE

CEBM

UNIVERSITY OF OXFORD

Critical Appraisal Skills Programme

Critical Appraisal Skills Programme (CASP)

Making sense of evidence

RAMMbo validity check

Representative: who did the subjects represent?

Allocation**:** randomised? Were groups similar at the start?

Maintenance**:** Were the groups treated equally? Were as many patients as possible followed-up?

Measurements

blinded or

objective

CEBM Oxford

Example…

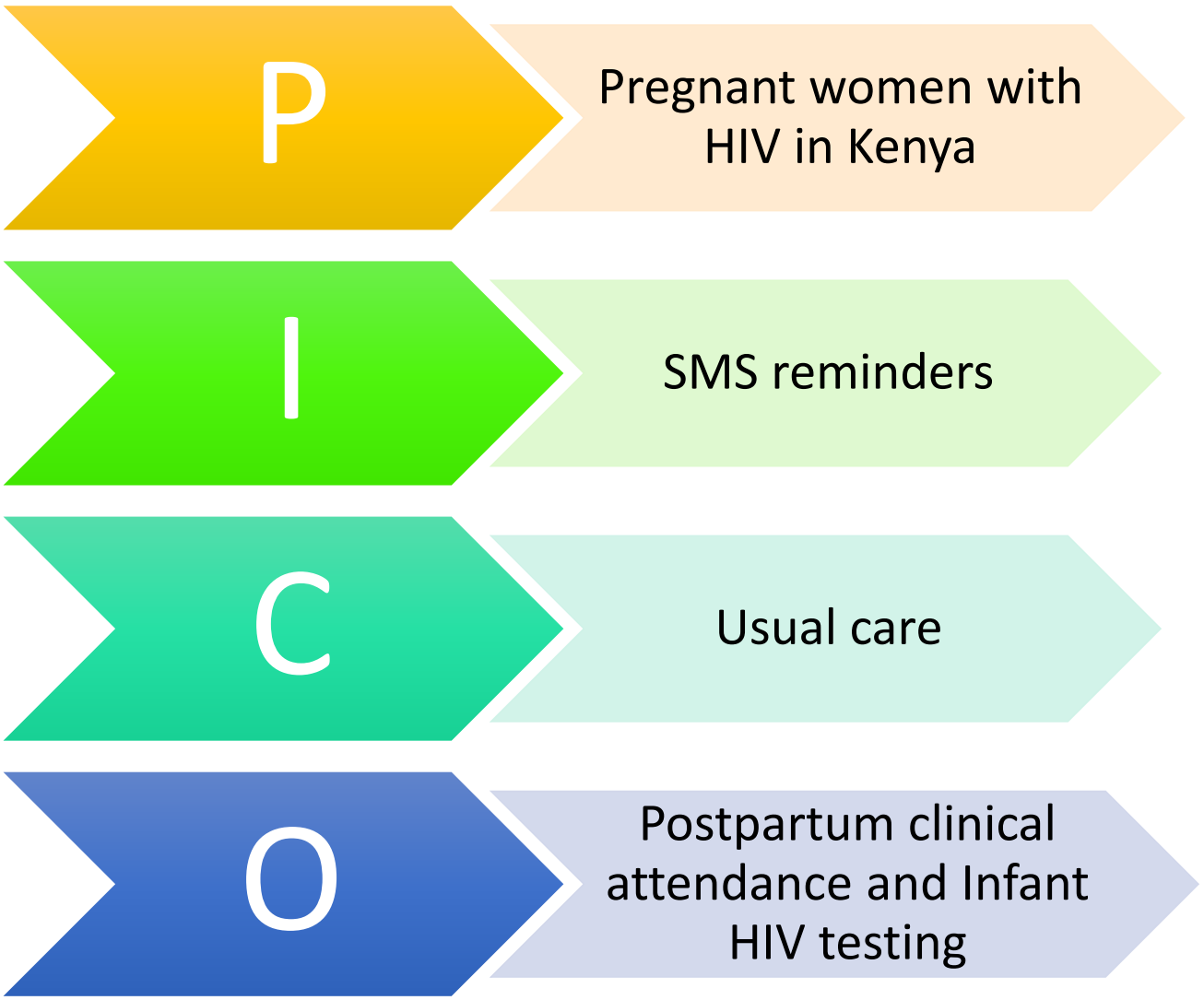
Texting improves testing: a randomized trial of two-way SMS to increase postpartum prevention of mother-to-child transmission retention and infant HIV testing

Thomas A. Odeny[a,b], Elizabeth A. Bukusi[a,c,d], Craig R. Cohen[e], Krista Yuhas[c], Carol S. Camlin[e] and R. Scott McClelland[b,c,f,g]

Objective: Many sub-Saharan African countries report high postpartum loss to follow-up of mother–baby pairs. We aimed to determine whether interactive text messages improved rates of clinic attendance and early infant HIV testing in the Nyanza region of Kenya.

Design: Parallel-group, unblinded, randomized controlled trial.

Odeny, et al 2014 AIDS


P
Pregnant women with HIV in Kenya
I
SMS reminders
C
Usual care
O
Postpartum clinical attendance and Infant HIV testing

RAMMbo validity check

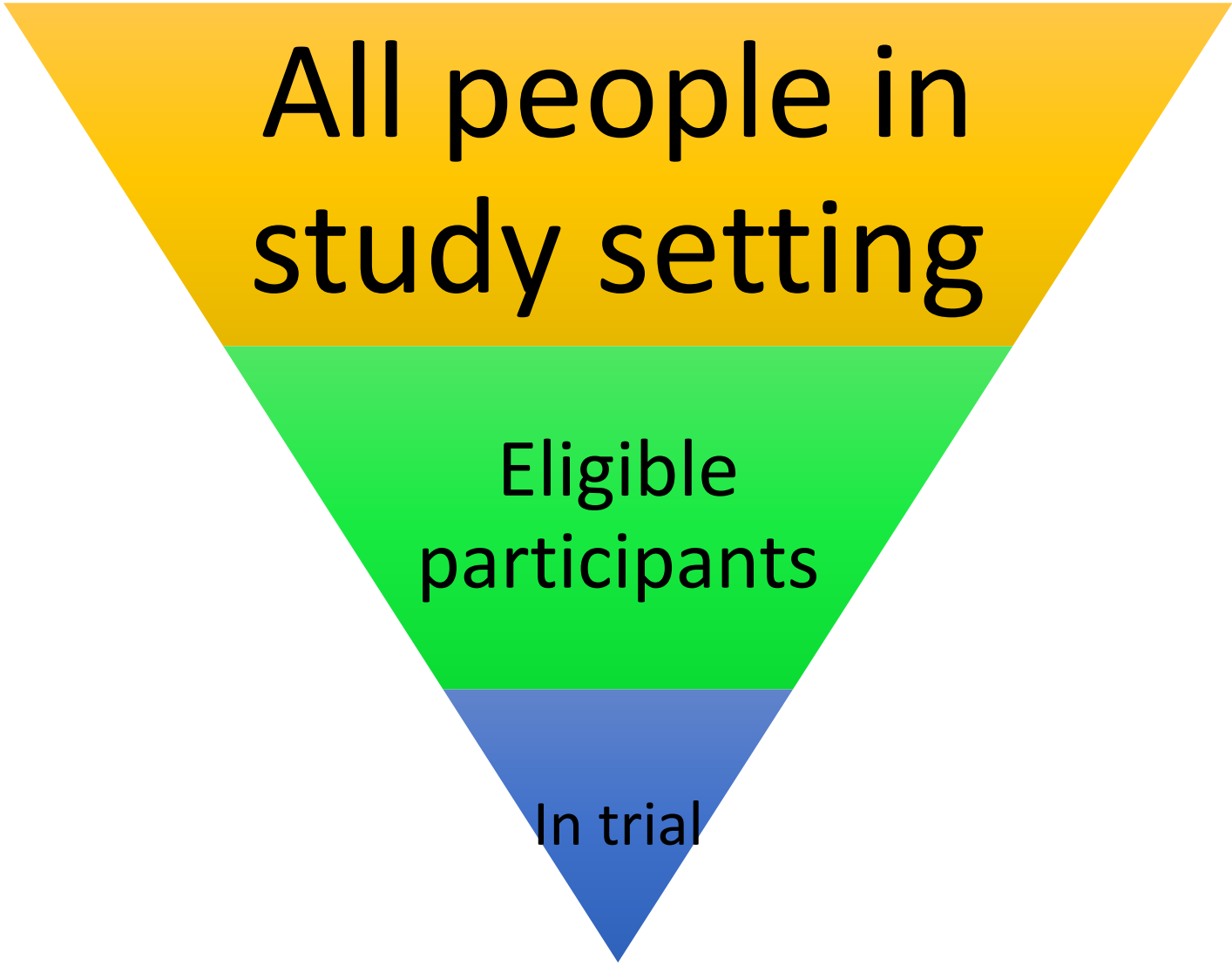
Representative: who did the subjects represent?

Allocation**:** randomised? Were groups similar at the start?

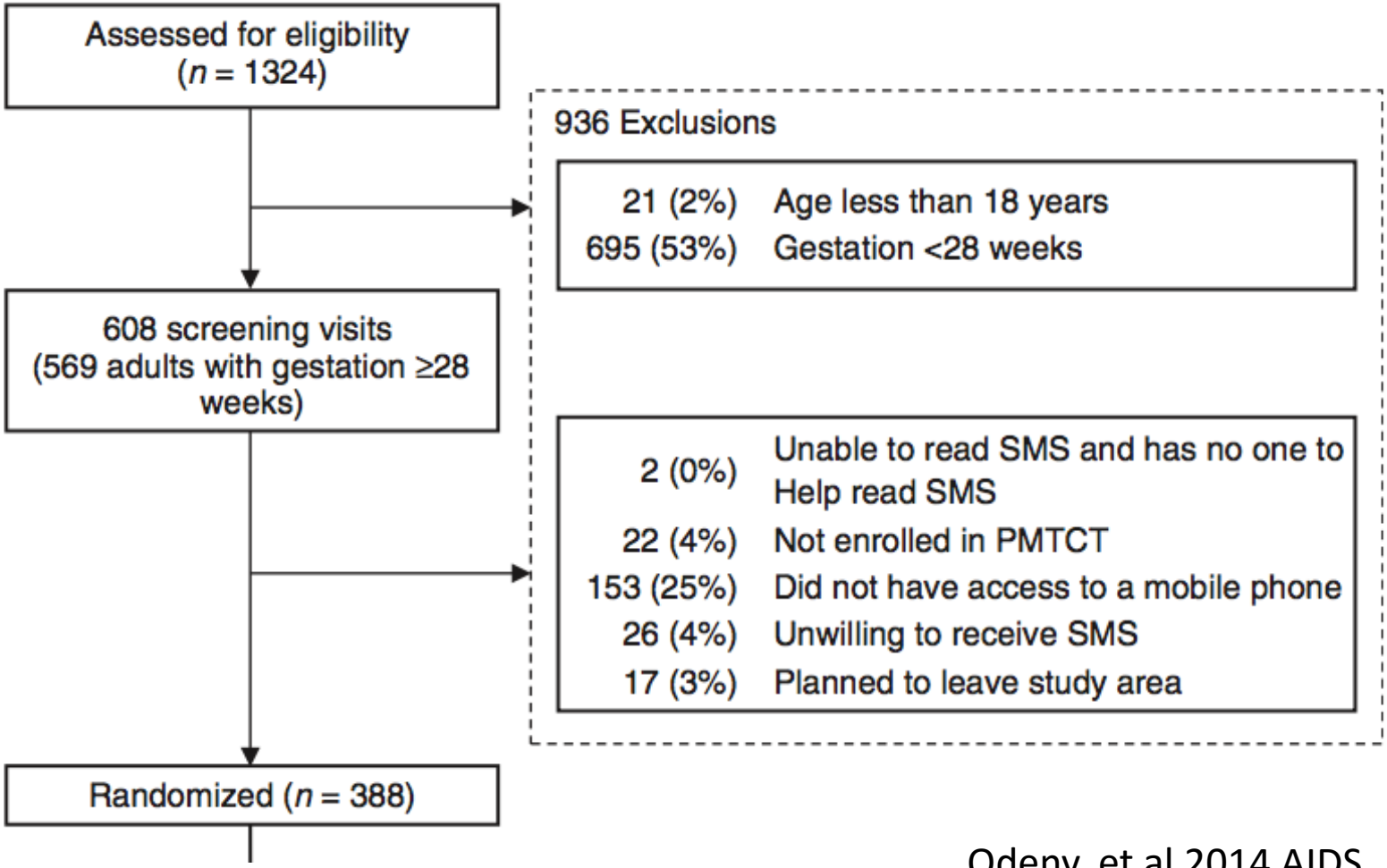
Maintenance**:** Were the groups treated equally? Were as many patients as possible followed-up?

Measurements **b**linded **or** **o**bjective

Representative: Are the trial subjects representative of patients in this setting?



Are the trial subjects representative of HIV positive pregnant women in this setting?



Odeny, et al 2014 AIDS

RAMMbo validity check

Representative: who did the subjects represent?

Allocation**:** randomised? Were groups similar at the start?

Maintenance**:** Were the groups treated equally? Were as many patients as possible followed-up?

Measurements **b**linded **or** **o**bjective

Why randomise?

- Minimises measured and unmeasured confounding

Minimising allocation bias

- Centralised computer randomisation the best
- Other methods such as sealed envelopes doubtful
- Non randomised: date of birth, alternate patients alternate days, etc

Allocation: How were participants randomised?

"A block randomization scheme with variable block sizes was used. Investigators and study staff were unaware of block numbers, sizes, or sequences. Intervention groups were assigned using sealed, opaque envelopes."

Odeny, et al 2014 AIDS

4. Were the groups similar at the start of the trial? ❑ Yes ❑ Can't tell ❑ No

Consider: Look at

- Other factors that might affect the outcome such as age, sex, social class, these may be called baseline characteristics

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net

Allocation: were the groups similar at the start?

Table 1. Maternal baseline characteristics.

Characteristics	SMS group ($n = 195$) n (%)	Control group ($n = 193$) n (%)
Maternal age		
18–24		65 (33.7)
25–34		111 (57.5)
35+		17 (8.8)
Gestational age at enrolment – median weeks (IQR)		34 (32–36)
Shared phone		50 (25.9)
Employed		39 (20.2)
Education		
None		3 (1.6)
Primary		110 (57.0)
Secondary		55 (28.5)
Post-secondary		25 (13.0)
Ethnicity		
Luo		177 (91.7)
Other		16 (8.3)
Married or with regular live-in partner (v	.7)	171 (88.6)
First pregnancy	(13.8)	29 (15.0)
WHO stage (highest recorded)		
1	110 (56.4)	103 (53.4)
2	55 (28.2)	57 (29.5)
3	23 (11.8)	27 (14.0)
4	7 (3.6)	6 (3.1)
Most recent $CD4^+$ cell cou		
<200	22 (11.3)	18 (9.3)
200–349	40 (20.5)	38 (19.7)
350–500	54 (27.7)	55 (28.5)
500+	78 (40)	82 (42.5)
On ART for own health	101 (51.8)	102 (52.8)
Received ZDV prophylaxis	85 (43.6)	81 (42.0)
Received ZDV + 3TC + NVP (delivery pac	60 (30.8)	53 (27.5)
Received ZDV + 3TC (post-delivery pack)	60 (30.8)	51 (26.4)
Nevirapine prophylaxis for baby issued	139 (71.3)	133 (68.9)
HIV diagnosed today	5 (2.6)	5 (2.6)
HIV counselling done with partner	40 (20.5)	49 (25.4)

3TC, lamivudine; IQR, interquartile range; NVP, nevirapine; ZDV, zidovudine.

No major differences between groups

Odeny, et al 2014 AIDS

RAMMbo validity check

Representative: who did the subjects represent?

Allocation**:** randomised? Were groups similar at the start?

Maintenance**:** Were the groups treated equally? Were as many patients as possible followed-up?

Measurements **b**linded **or** **o**bjective

Maintenance: Were the groups treated equally?

5. Aside from the experimental intervention, were the groups treated equally? ☐ Yes ☐ Can't tell ☐ No

Study staff called participants in the SMS arm weekly beginning at 38 weeks gestation to ascertain whether delivery had occurred. Delivery dates for participants in the control arm were abstracted from clinic records. If control women did not return, they were contacted either in person or by phone. Women's return visits and infant HIV testing data were extracted from clinic records.

Odeny, et al 2014 AIDS

6. Were all of the patients who entered the trial properly accounted for at its conclusion?

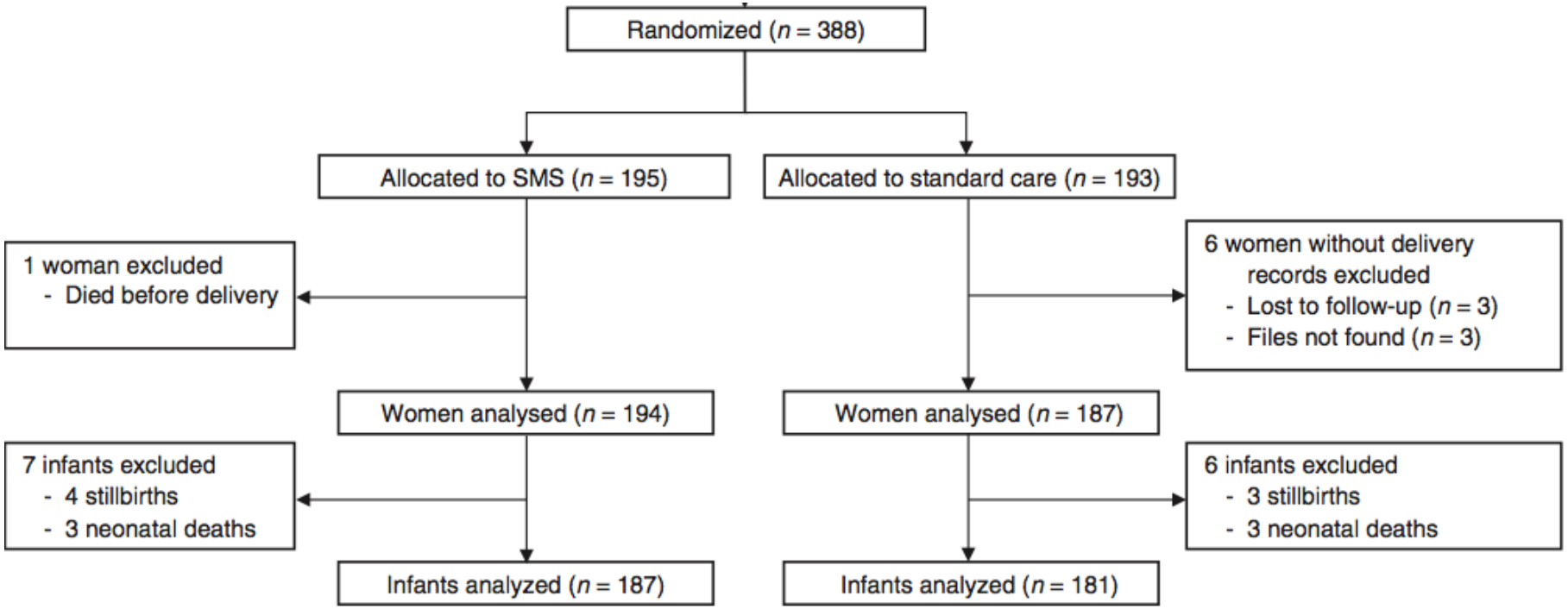
❑ Yes ❑ Can't tell ❑ No

Consider:

- Was the trial stopped early?
- Were patients analysed in the groups to which they were randomised?

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net

Maintenance: were as many patients as possible followed-up?



1. Trial profile showing flow of study participants.

Odeny, et al 2014 AIDS

Intention to treat

- Once a participant is randomised, they should be analysed to the group they were assigned to
- Pros
  - Reflects "real life" e.g non compliance
  - Unbiased estimate of true effect
  - Maintains sample size thus maintaining statistical power
- Cons
  - Noncompliance provides little data on efficacy
  - Treatment effect may be conservative
  - Dropouts/non-compliant/compliant subjects are different

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/

RAMMbo validity check

Representative: who did the subjects represent?

Allocation**:** randomised? Were groups similar at the start?

Maintenance**:** Were the groups treated equally? Were as many patients as possible followed-up?

Measurements **b**linded or **o**bjective

Detailed questions

3. Were patients, health workers and study personnel blinded?

❑ Yes ❑ Can't tell ❑ No

Consider:

- Health workers could be; clinicians, nurses etc
- Study personnel – especially outcome assessors

Measurements blinded or objective

Women's return visits and infant HIV testing data were extracted from clinic records.

Odeny, et al 2014 AIDS

(B) What are the results?

7. How large was the treatment effect?

Consider:

- What outcomes were measured?
- Is the primary outcome clearly specified?
- What results were found for each outcome?
- Is there evidence of selective reporting of outcomes?

8. How precise was the estimate of the treatment effect?

Consider:

- What are the confidence limits?
- Were they statistically significant?

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net

What does this study tell us?

- **P values** (hypothesis testing):
  - Tests to exclude the null hypothesis
- **Confidence intervals** (estimation of effect)
  - Range of values within which the true effect is likely to lie
  - Wider the confidence interval, less precision in result
- **Relative Risk**
- **Absolute Risk**
- **Odds Ratios**
- **Number needed to treat**

In the SMS group, 38 of 194 (19.6%) women attended a postpartum clinic visit compared to 22 of 187 (11.8%) in the control group [relative risk (RR) 1.66, 95% confidence interval (CI) 1.02–2.70, $P = 0.04$].

In the per-protocol analysis, women in the SMS arm had a significantly higher probability of attending clinic within 8 weeks compared to those in the control arm (RR 1.83, 95% CI 1.11–3.01).

Odeny, et al 2014 AIDS

Odeny, et al 2014 AIDS

(C) Will the results help locally?

9. Can the results be applied in your context?

(or to the local population?)

☐ Yes ☐ Can't tell ☐ No

Consider:

- Do you have reason to believe that your population of interest is different to that in the trial
- If so, in what way?

10. Were all clinically important outcomes considered? ❑ Yes ❑ Can't tell ❑ No

Consider:

- Is there other information you would like to have seen?
- Was the need for this trial clearly described?

---

11. Are the benefits worth the harms and costs? ❑ Yes ❑ Can't tell ❑ No

Consider:

- Even if this is not addressed by the trial, what do you think?

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net

Conclusion

- Critical appraisal helps us decide whether evidence is valid, what the results tell us and whether the study is relevant to our setting
- Checklists are available to help
- Don’t believe everything you read in journals!